K.P. Pylypenko

# Identification of the probability density of the sum of the signal with gaussian mixture distribution and gaussian white noise

**В работе решена задача получения плотности вероятностей процесса в виде суммы белого шума и сигнала, плотность вероятностей которого описывается смесью нормальных распределений. Предложен метод восстановления плотности вероятностей полезного сигнала с использованием оценок кумулянтных коэффициентов и алгоритма идентификации.**

**The problem of obtaining the probability density of a process that is a sum of white noise and a signal with a Gaussian mixture distribution is solved in this paper. Using estimates $\sigma^2$, $\gamma_4$ and $\gamma_6$, and the identification algorithm the probability density of the payload signal was reconstructed.**

**Keywords**: *Gaussian mixture of distributions, probability density, cumulant coefficients, white noise.*

## Introduction

The problem of analyzing a process comprising signal and noise is common in engineering practice and is most often encountered when solving optimal signal reception tasks [1]. Here, processing of the received signals is typically based on the methods of mathematical statistics. A classic example of an optimum receiver is a receiver of known (deterministic) signals in the presence of white Gaussian noise.

The main difficulties arise when the signal is not deterministic, but stochastic, particularly if it is necessary not only to detect the signal, but also to determine its characteristics and estimate its parameters [2]. Detecting and estimating the characteristics of fluctuation signals in the presence of noise is one of the applications. It should be noted, that the quality of the solution of this problem depends on the correct choice of signal and noise models. One of the models that describe the distribution of a fluctuation signal is a mixture of distributions [3], in [4], a mixture of distributions is used to describe the probability density of a speech signal.

In this paper we propose a solution of the problem of obtaining the probability density of the process that is the sum of white Gaussian noise and signal with a Gaussian mixture distribution, i.e.

$$\zeta(t) = s(t) + \xi(t),\qquad (1)$$

where $s(t)$ is the signal, $\xi(t)$ is Gaussian white noise.

Also taken into consideration cumulant coefficients of the process (1) and their usage for solving problems of detection and recover probability density of payload signals.

To obtain the probability density of the process $\xi(t)$ we mast use the convolution formula

$$p_\zeta(y) = \int\limits_{-\infty}^{\infty} p_\xi(y - x) p_s(x) dx,\qquad (2)$$

which describes relation between probability densities of each process and sum of that processes.

As noise $\xi(t)$, we will consider a Gaussian process with zero mean, its probability density being described by the following expression:

$$p_\xi(x) = \frac{1}{\sigma_\xi \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_\xi^2}}.\qquad (3)$$

Direct determination of the probability density (2) is not always a trivial task, and, in many cases, obtaining an analytical expression is not possible.

## The probability density

Let us consider the case where the probability density can be obtained directly from the expression (2).

Let the distribution of the process $s(t)$ be described by a mixture of distributions

$$p_s(x) = \frac{d_1}{\sigma_1} \varphi\left(\frac{x - m_1}{\sigma_1}\right) + \frac{d_2}{\sigma_2} \varphi\left(\frac{x - m_2}{\sigma_2}\right),\qquad (4)$$

where the coefficients $d_1$ and $d_2$ meet the conditions:

$$d_1 + d_2 = 1;\ d_1 > 0, d_2 > 0,\qquad (5)$$

$\varphi(x)$ is the probability density of a standard Gaussian random variable

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Let us find the probability density of the sum of the process (4) and Gaussian noise (3) with using expression (2). Integral in (2) becomes

$$p_\zeta(y) = \int_{-\infty}^{\infty} p_\xi(y-x)p_s(x)dx =$$

$$= \int_{-\infty}^{\infty}\left(\frac{d_1}{\sigma_1}\varphi\left(\frac{x-m_1}{\sigma_1}\right) + \frac{d_2}{\sigma_2}\varphi\left(\frac{x-m_2}{\sigma_2}\right)\right) \times$$

$$\times \frac{1}{\sigma_\xi}\varphi\left(\frac{y-x}{\sigma_\xi}\right)dx = \frac{d_1}{\sqrt{\sigma_1^2+\sigma_\xi^2}}\varphi\left(\frac{y-m_1}{\sqrt{\sigma_1^2+\sigma_\xi^2}}\right) +$$

$$+ \frac{d_2}{\sqrt{\sigma_2^2+\sigma_\xi^2}}\varphi\left(\frac{y-m_2}{\sqrt{\sigma_2^2+\sigma_\xi^2}}\right).$$

(6)

From expression (6) we can conclude, that obtained probability density is a mixture of Gaussian distributions, and the only difference betweeen (4) to (6) is the fact that the noise variances $\sigma_\xi^2$ was added to the variances of the components in (4).

Now we generalize the expression (6). Let the probability density of a process $s(t)$ be represented as a mixture of distributions with an unlimited number of components

$$p(x) = \sum_{l=1}^{\infty}\frac{d_l}{\sigma_l}\varphi\left(\frac{x-m_l}{\sigma_l}\right),$$

(7)

where $\sum_{l=1}^{\infty} d_l = 1$, $d_l > 0$.

Then the probability density of the sum of this process and Gaussian noise, by analogy with (6), is defined by

$$p(x) = \sum_{l=1}^{\infty}\frac{d_l}{\sqrt{\sigma_l^2+\sigma_\xi^2}}\varphi\left(\frac{x-m_l}{\sqrt{\sigma_l^2+\sigma_\xi^2}}\right).$$

(8)

Thus, if the probability density of the process is described by a mixture of distributions, then, using expression (8), we can always obtain an expression for the probability density of this process with additive Gaussian noise.

Now we consider one of the most important case of Gaussian mixture of distribution which is unimodal two-component Gaussian mixture. Probability density of that mixture of distribution in this case can be obtained from (4) by letting $m_1$ and $m_2$ be equal to zero:

$$p_s(x) = \frac{d_1}{\sigma_1}\varphi\left(\frac{x}{\sigma_1}\right) + \frac{d_2}{\sigma_2}\varphi\left(\frac{x}{\sigma_2}\right).$$

(9)

It is easy to obtain the mean, variance and cumulant coefficients of such a mixture [5]:

$$m = 0, \sigma^2 = d_1\sigma_1^2 + d_2\sigma_2^2,$$

$$\gamma_3 = 0,$$

$$\gamma_4 = \frac{3\left(d_1\sigma_1^4 + d_2\sigma_2^4 - \left(d_1\sigma_1^2 + d_2\sigma_2^2\right)^2\right)}{\left(d_1\sigma_1^2 + d_2\sigma_2^2\right)^2},$$

$$\gamma_5 = 0,$$

(10)

$$\gamma_6 = 15\left[d_1\sigma_1^6 + d_2\sigma_2^6 + 2\left(d_1\sigma_1^2 + d_2\sigma_2^2\right)^3 - \right.$$

$$\left. -3\left(d_1\sigma_1^2 + d_2\sigma_2^2\right)\left(d_1\sigma_1^4 + d_2\sigma_2^4\right)\right]\bigg/\left(d_1\sigma_1^2 + d_2\sigma_2^2\right)^3.$$

In purpose of illustration let the parameters of distribution (9) be defined as:

$$m_1 = m_2 = 0,$$

$$\sigma_1 = 1, \sigma_2 = 4,$$

$$d_1 = 0,6, d_2 = 0,4.$$

(11)

Fig. 1 depicts the plots of normalized probability densities of the payload signal and the noised signal with signal-to-noise ratio $\left(SNR = \sigma_s^2 / \sigma_\xi^2\right)$ values 10, 1, 0.5 and 0.1. Fig. 1 illustrates that with decreasing SNR the distribution of the process tends to a Gaussian distribution.

**Cumulant coefficients**

Let us consider the cumulant coefficients of the process (1). Since the probability density can be uniquely defined by infinite series of moments [6], cumulants, and, consequently, the cumulant coefficients, knowing the cumulant coefficients allows us to obtain probability density in cases where it is not possible analytically, as well as estimate the probability density experimentally using momentum representation.

The cumulants of the distribution (2) are given by the formula [6]

$$\kappa_k^\zeta = j^{-k}\left[\frac{d^k \ln f_\zeta(u)}{du^k}\right]\bigg|u = 0,$$

(12)

where $f(u)$ is the characteristic function, which is equal to

$$f_\zeta(u) = \int_{-\infty}^{\infty} e^{jux}p_\zeta(x)dx.$$

(13)

According to the properties of the characteristic function, $f_\xi(u)$ can be written as
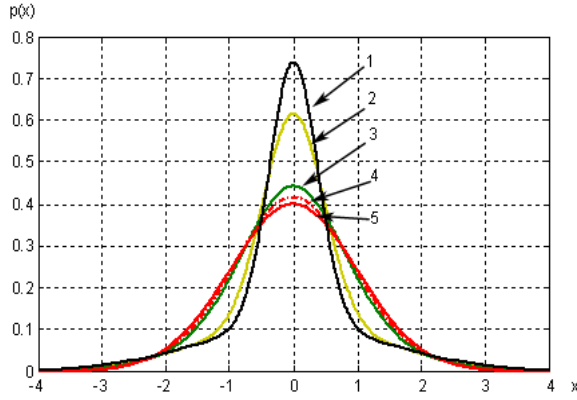
$$f_\zeta(u) = f_s(u)f_\xi(u),$$

(14)

**Fig. 1. PDFs of the payload and the noised signals:**

**1 – payload signal; 2 – SNR=10; 3 – SNR=1; 4 – SNR=0,5; 5 – SNR=0,1** and the expression for the cumulant coefficients (12) becomes

$$\kappa_k^\zeta = j^{-k}\left[\frac{d^k \ln\left(f_s(u)f_\xi(u)\right)}{du^k}\right]\Bigg|u = 0 =$$

$$= j^{-k}\left[\frac{d^k \ln f_s(u)}{du^k} + \frac{d^k \ln f_\xi(u)}{du^k}\right]\Bigg|u = 0 =$$

$$= \kappa_k^s + \kappa_k^\xi. \tag{15}$$

Knowing the cumulants of the distribution $p_s(x)$ and $p_\xi(x)$, we can easily obtain the cumulants of the process, which is a sum of signal $s(t)$ and noise $\xi(t)$.

The characteristic function in the case of a normal distribution (3) of $\xi(t)$ is equal to

$$f_\xi(u) = e^{-\frac{u^2\sigma_\xi^2}{2}}, \tag{16}$$

and cumulants $k_k^\xi$, according to (12),

$$\kappa_2^\xi = \sigma_\xi^2;$$
$$\kappa_k^\xi = 0, k = 3,4,... \tag{17}$$

Based on the expressions (15) and (17), the cumulants of the process (1) can be described by formulas

$$\kappa_2^\zeta = \kappa_2^s + \sigma_\xi^2;$$
$$\kappa_k^\zeta = \kappa_k^s, k = 3,4,... \tag{18}$$

The last expression indicates that the summation of process $s(t)$ with Gaussian noise $\xi(t)$ changes only the second cumulant, and all others remain unchanged.

If the signal-to-noise ration is given by the equation

$$SNR = \frac{\kappa_2^s}{\sigma_\xi^2}, \tag{19}$$

then, according to the expression (18), the second cumulant $k_k^\zeta$ is

$$\kappa_2^\zeta = \kappa_2^s\left(\frac{1+SNR}{SNR}\right). \tag{20}$$

To obtain expressions for the cumulant coefficients of the process (1) we shall use expressions (18) and (20) and the formula for cumulant coefficients [6]

$$\gamma_k^\zeta = \frac{\kappa_k^\zeta}{\left(\kappa_2^\zeta\right)^{\frac{k}{2}}} = \left(\frac{SNR}{1+SNR}\right)^{\frac{k}{2}} \times$$

$$\times \frac{\kappa_k^s}{\left(\kappa_2^s\right)^{\frac{k}{2}}} = \left(\frac{SNR}{1+SNR}\right)^{\frac{k}{2}} \gamma_k^s. \tag{21}$$

Based on the expression (21) we can conclude that adding Gaussian noise to the process multiplies cumulant coefficient by value $\left(\frac{SNR}{1+SNR}\right)^{\frac{k}{2}}$.

### Detection of the signal

Now let us investigate the possibility of detection a noise-like signal with probability density (9) in the presence of Gaussian noise as well as the possibility of reconstruction its probability density.

The algorithm of modeling of mixture of distributions (4) described in detail in [7], so in this paper we will not pay attention to it. As the parameters of the distribution of signal (9) we will take the parameters (11). Note that for all the examples the length of the realization of signal is $N = 10^6$.

In the Table 1 simulation results are presented. For each value of SNR using expressions (10) and (21) theoretical values variance $\sigma^2$ and cumulant coefficients $\gamma_4$ and $\gamma_6$ were obtained as well as their estimations $\overline{\sigma}^2$, $\overline{\gamma}_4$ and $\overline{\gamma}_6$ with usage of modeling of signals.

Using estimates $\overline{\sigma}^2$, $\overline{\gamma}_4$ and $\overline{\gamma}_6$, and the identification algorithm proposed in [8], we can reconstruct the probability density of the payload signal. For this purpose expression (21) was used to recalculate the cumulant coefficients and the variance for each SNR value. The theoretical probability density and the probability density of reconstracted signals with different SNR values are

compared in Fig. 2. As the reconstruction error $\delta$ we used the integral metric

$$\delta = \rho(p, p_0) = \int_{-\infty}^{\infty} |p(x) - p_0(x)| dx$$

,

where $p_0(x)$ is the theoretical probability density.

Fig. 2 and Table. 2 show that if the SNR value is greater than 0,1 (-20 dB) then reconstructed probability density almost fully matches to the theoretical, only when the value of SNR is 0.1 the result of the identification can be unsatisfactory.
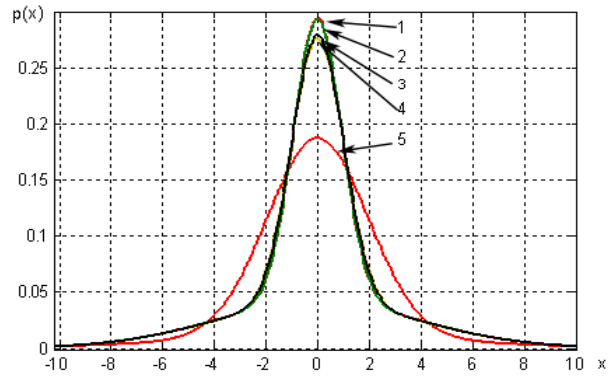


**Fig. 2. PDFs of the payload and reconstructed signals:**

**1 – SNR=0,5; 2 – SNR=1; 3 – payload signal; 4 – SNR=10; 5 – SNR=0,1**

**Table 1. Cumulant coefficients for different value of SNR**

| SNR | 0,1 | 0,5 | 1 | 10 | Without noise |
|---|---|---|---|---|---|
| $\sigma^2$ | 77 | 21 | 14 | 7,7 | 7 |
| $\overline{\sigma}^2$ | 77,0362 | 20,9935 | 13,9917 | 7,7274 | 6,9962 |
| $\gamma_4$ | 0,0273 | 0,3673 | 0,8265 | 2,7323 | 3,3061 |
| $\overline{\gamma}_4$ | 0,0328 | 0,3706 | 0,8413 | 2,7553 | 3,3284 |
| $\gamma_6$ | 0,0053 | 0,2624 | 0,8856 | 5,3227 | 7,0846 |
| $\overline{\gamma}_6$ | 0,0361 | 0,2416 | 0,8617 | 5,6460 | 7,3995 |

**Table 2. Reconstruction errors**

| SNR | 0,1 | 0,5 | 1 | 10 |
|---|---|---|---|---|
| $\delta$ | 0,3338 | 0,0399 | 0,0347 | 0,0087 |

## Conclusion

In this paper a general analytical expression for the probability density of a process that comprises a sum of a signal and noise, where the signal's distribution is a mixture of distributions, and the noise is a Gaussian process, was obtained.

Analysis of cumulant coefficients of the process demonstrated that the cumulant coefficients of the process (1) and the cumulant coefficients of the payload signal differ by value $\left(\dfrac{SNR}{1+SNR}\right)^{\frac{k}{2}}$. This fact allows us to solve the problem of detecting a payload signal in the presence of Gaussian noise and to identify its probability density with a mixture of distributions.

The possibility of detecting the payload signal and acquiring its characteristics for different signal-to-noise ratios was demonstrated with numerical modeling.

## Literature

1. *Тихонов В. И.* Оптимальный приём сигналов. – М.: Радио и связь, 1983. – 320 с.
2. *Куликов Е. И.*, Трифонов А. П. Оценка параметров сигналов на фоне помех. М.: Советское радио, 1978. – 296 с.
3. *Красильников А.И.*, Пилипенко К.П. Идентификация плотности вероятностей пуассоновской последовательности прямоугольных импульсов с гауссовским распределением амплитуд // Электроника и связь. Тематический выпуск «Электроника и нанотехнологии». – 2011. – № 61(2). – С. 68–73.
4. *Величкин А.И.* Передача аналоговых сообщений по цифровым каналам связи. – М.: Радио и связь, 1983. – 240 с.
5. *Тихонов В.И.* Статистическая радиотехника. – 2-е изд., перераб и доп. – М.: Радио и связь, 1982. – 624 с.

6.  *Малахов А.Н.* Кумулянтный анализ случайных негауссовских процессов и их преобразований. – М.: Сов. радио, 1978. – 376 с.

7.  *Красильников А.И.*, Пилипенко К.П. Моделирование дискретных смесей распределений // Электроника и связь. – 2010. – № 55. – С. 57–61.

8.  *Красильников А.И.*, Пилипенко К.П. Применение двухкомпонентной гауссовской смеси для идентификации одновершинных симметричных плотностей вероятностей // Электроника и связь. – 2008. – № 5(46). – С. 20–29.

*Национальный технический университет Украины «Киевский политехнический институт»*